2023-07

Working paper. Economics

ISSN 2340-5031

OPTIMAL CONTRACTS WHEN THE PLAYERS THINK DIFFERENT*

MARTIN DUMAV, URMEE KHAN and LUCA RIGOTTI

uc3m | Universidad Carlos III de Madrid

Optimal contracts when the players think different*

Martin Dumav[†] Urmee Khan[‡] Luca Rigotti[§]

September 25, 2023

Abstract

In a moral hazard model with heterogeneous beliefs, we show that the efficient risk-sharing contract does not result in a constant wage and the optimal first-best contract may not be increasing in output. When actions are unobservable, heterogeneity in beliefs implies that the monotone likelihood ratio ranking does not ensure that the wage scheme in the optimal contract is non-decreasing in output. This is because differences in beliefs may affect the incentive provision in a non-monotone way. The standard monotonicity result with common beliefs extends to belief heterogeneity when the agent is more optimistic than the principal. Yet, in the reverse case, the optimal contract can be non-monotone.

Keywords: Contracting, heterogeneous beliefs, monotone likelihood ratio, moral hazard

JEL Classification: D82, D86, M52

*Many thanks to Antonio Cabrales, Philipp Denter, Angel Hernando, Jan Knoepfle, Roberto Serrano, William Fuchs, seminar participants at Universidad Carlos III de Madrid, Uni Pompeau Fabra, European University Institute, Uni Alicante. Dumav gratefully acknowledges support from from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie Grant # 896463, and the Ministerio Economia y Competitividad (Spain) through grants ECO2017-86261-P and MDM 2014-0431.

[†]Universidad Carlos III de Madrid

[‡]University of California Riverside

[§]University of Pittsburgh

1 Introduction

Since the classic work of Holmström (1979), moral hazard theory has focused on the interplay between differences in information and differences in risk-aversion of two players, the principal and the agent that share common beliefs. However, principal and agent might differ in another dimension: their beliefs. Consumer surveys clearly indicate differences in beliefs among agents (see, e.g., Fan et al. (2020)). Many real-life contracts involve individuals with very different expectations about their performance. Beliefs can differ for many reasons; for example, principal and agent may not share a common understanding of the production technology, due to possibly having diverse experiences in the past, or having information coming from different or conflicting sources. In this paper we extend the canonical moral hazard model to allow for such heterogeneity. In particular, we model heterogeneity in beliefs about the output levels resulting from agents' action choices.

We study a generalization of the model in Grossman and Hart (1983) (henceforth GH), where each of the agent's actions is associated with a pair of probability distributions over output levels, one perceived by the agent and another by the principal. Except for the divergence of beliefs, the set-up follows closely that of the traditional model. The principal is risk-neutral, the agent is risk-averse, the contract stipulates a payment from the principal to the agent, i.e. a wage, that depends on the realized output level, and the parties' beliefs are common knowledge. Indeed, a version of our model where for *all* actions the principal and the agent share common beliefs specializes to the canonical model. We use this model to understand the role belief heterogeneity plays in the fundamental efficiency-incentives trade-off, and to what extent it affects the qualitative properties of optimal contracts, in particular, their monotonicity.

We begin by considering the perfect information case, where the agent's action is observed by the principal and can be specified in the contract. With homogeneous beliefs, optimal risk-sharing dictates that the agent receives a constant wage. This result is a straightforward implication of Borch (1962) rule for optimal risk-sharing with common beliefs since, in the absence of an incentive problem, the risk-neutral party absorbs all the risk. With heterogeneous beliefs the Borch rule implies that a constant contract is no longer efficient because the contract must now account for differences in beliefs in addition to the agent's risk-aversion.

The first-best contract in an environment with heterogeneous beliefs has some more interesting features: First, wages are not necessarily monotone in output because they must account for the *relative* difference in beliefs between the two parties. The relative "optimism" or "pessimism" over a particular output level dictates whether the agent will be paid higher or lower wages for that output compared to another. For example, if the agent places higher probability on some output it could be cheaper to pay him a higher wage for that output and a lower wage for another output level over which the principal has more optimistic beliefs. Thus, the principal can exploit the difference in beliefs and end up with a lower expected wage bill than paying a constant wage across all output realizations.

Second, following the same intuition, our comparative static results show that in the first-best scenario wages move in the opposite direction of the (small) change in the principal's belief. Finally, even for the first-best it is hard to determine the entire shape of the wage function in general given that wages are sensitive to belief differences. To get a sharper characterization, we examine the case where the beliefs of the two parties over output levels can be ranked by the Monotone Likelihood Ratio Property (MLRP). We show that if the principal is more optimistic than the agent (in MLRP ranking), then cost minimizing wages are *decreasing*, while they are *increasing* in the opposite scenario.

When actions are not observable, the belief structure involves multiple dimensions of comparison and the analysis of the optimal contract is more complex. For example, the beliefs of the agent over multiple actions play into the incentive constraints, while for a particular action the two players have different beliefs. Hence, in this case the incentive problem of the agent cannot be studied in isolation from the principal's beliefs. These elements are reflected in the first-order conditions of the problem that entail different components which depend on the beliefs of different parties: one term depends on the beliefs of the principal, while the other two depends only on the beliefs of the agent. As in the standard case, the optimal wage scheme involves a term capturing the risk-sharing role and a term capturing the incentive provision role. Unlike the standard case, the incentive provision term can sometimes become irrelevant, as it may be undermined by the principal's ability to exploit the differences in beliefs to provide cheaper incentives.

Because of this more complicated trade-off between risk-sharing and incentives, for tractability, we focus on the model with two actions that serves as a base for the core results of our work: In the common beliefs case, monotonicity follows from the MLRP ranking of beliefs for different actions. In our model,we show that monotonicity features of the optimal wage scheme can be driven by the difference between the beliefs of the two parties. Contrary to the standard framework of GH, the MLRP ranking of the parties' beliefs does not ensure that the optimal wage scheme is non-decreasing in output when beliefs are heterogeneous. We show that under MLRP condition the monotonicity result of GH extends to belief heterogeneity when the agent is more optimistic than the principal. On the other hand, when the agent is relatively more pessimistic the optimal contract can be non-monotone.

Our analysis provides also novel insights into the marginal effects of disagreement on incentives. Using a semi-parametric setting with two actions and three output levels for tractability, we show that small perturbations in the principal's relative optimism drive the wage in the optimal contract in the opposite direction. To the best of our knowledge, there is no general characterization of the responsiveness of wage schemes to perturbations in beliefs in the existing literature. For that reason, we view our result as novel.

De la Rosa (2011) also focuses on characterizing the shape of the optimal contract, but offers a particular formulation of heterogeneous beliefs.[1]. With observable actions, our results generalize those in De la Rosa (2011). With unobservable actions, the assumed heterogeneity is different and results are not directly comparable, although they are based on similar intuitions. Also, our work is related to the literature in finance that analyzes the consequences of managerial overconfidence. Gervais et al. (2011), for example, study the capital budgeting decisions of an overconfident manager. This approach differs from ours, as we do not explicitly focus on the agent's overconfidence and, following GH, pursue a more general moral hazard setting. Finally, characterizing systematically the cost of implementation in relation to the agent´s relative optimism/pessimism is a difficult problem to tackle in general. For that reason, we restrict our analysis to two actions and finite output levels and concentrate in characterizing the monotonicity of the contract. To exemplify how one can get misled in this framework, we show that the MLRP ranking in the case of two actions does not guarantee monotonicity of the wage contract when beliefs are heterogeneous. Santos-Pinto (2008) proposes to analyze the consequences of overconfidence from the part of the agent on the principal's welfare when the wage scheme is increasing with output for a finite number of actions and finite number of output levels. Although our framework is more restrictive, our conclusions result from a rigorous analysis that does not make use of local approximations and shows that the direction of the asymmetry in beliefs can affect the nature of the optimal contract. For example, we show that the results of Santos-Pinto (2008) may hold only in the special case of overoptimism from the part of the agent and cannot extend to the case of overpessimism.

Our paper also relates to the literature on robustness of contracts under ambiguity. It turns out that our setup of a moral hazard problem with probabilistic but heterogeneous beliefs can be viewed as falling somewhere in between the standard homogeneous beliefs models and models with ambiguity as sets of probabilities. Robust contracts in the latter setting turn out to have simple forms: e.g. linear (Dumav and Khan (2021)) or step functions (Lopomo et al. (2011)). Interestingly, the disagreement between the parties is what drives these shapes, in

[1] Relative to our work, De la Rosa (2011) makes specific parametric assumptions about the probability distributions, whereas we do not impose any such restrictions. Also, the notion of heterogeneity considered in that paper depends on the optimal contract, while it is fully exogenous in our case.

particular these are the contracts that either eliminate the disagreement, or exploit it, for Pareto improvement. Disagreement also drives the shape of the contracts in our model, but the contracts themselves turn out to be more complicated and generally sensitive to the details of the belief structure. These differences in properties of optimal contracts arise because the disagreement in our model is exogenously fixed, whereas the disagreement in models of moral hazard with ambiguity are endogenous to the contractual form. Hence in the latter setting, the contracts themselves become the tool to shape the equilibrium level of disagreement, which is not possible in our setting given exogenous heterogeneity in beliefs.

The reminder of the paper is organized as follows: the next section describes the model and Section 3 presents the observable action case. Section 4 presents results under unobservable actions and Section 5 concludes. The appendix collects various proofs of propositions and lemmas presented in the main text.

2 Model

We use the classical moral hazard environment of GH. The principal owns a technology that produces stochastic output with the agent's effort which determines the probability distribution over outputs. Formally, there are a finite number S of output levels, the set of outputs is denoted by $\mathcal{Y} =: \{y_1, \ldots, y_S\}$ and we label output levels s so that higher subindexes correspond to higher outputs: $0 \leq y_1 < y_2 < ... < y_S$ with the highest output y_S being finite. In referring to different output levels it will be helpful to denote the state space by $\mathcal{S} = \{1, \ldots, S\}$.

The principal chooses the wages paid to the agent contingent on the output, denoted by the vector $w \in \mathbf{R}^S$, and keeps the difference between output and wages for herself. The agent chooses an action $a \in \mathcal{A}$, which is a compact set. In keeping with the traditional moral hazard model, we assume the principal is risk-neutral while the agent is risk-averse. More specifically, we make the following assumptions on the preferences:

Assumption 1. The agent's utility function is additively separable in monetary outcomes and effort $U(w, a) = u(w) - g(a)$. Here the utility function u is continuous, strictly increasing, and concave function, which satisfies an Inada condition, $\lim_{w \searrow \underline{W}} u(w) = -\infty$ for some $\underline{W}$. The effort cost function $g(a)$ is continuous and increasing in a. The principal is risk-neutral with a linear utility function.

As a novel element in the model of moral hazard, we allow the principal and agent to have different beliefs about how each action influences the distribution of output. Formally, for each action taken by the agent there are two probability distributions, $\pi^P(a) \in \Delta(\mathcal{Y})$ and

$\pi^A(a) \in \Delta(\mathcal{Y})$, representing the beliefs of the principal and the agent respectively ($\Delta(\mathcal{Y})$ denotes the S-dimensional simplex). We assume that the probability distributions $\pi^P(a)$ and $\pi^P(a)$ are continuous functions of a. When the agent chooses an action a the principal's utility function is given by

$$U^P(w;a) = \sum_{s=1}^{S} \pi_s^P(a)(y_s - w_s) \tag{2.1}$$

and the agent's utility function is given by

$$U^A(w;a) = \sum_{s=1}^{S} \pi_s^A(a)u(w_s) - g(a). \tag{2.2}$$

We assume that the parties' beliefs are common knowledge and have a common full support. More specifically, we make the following assumption on the belief structure.

Assumption 2. For each action $a \in \mathcal{A}$, the beliefs of the principal and the agent, $\pi^P(a)$ and $\pi^A(a)$, respectively, are common knowledge and have a common full support:

$$\text{For all } a \in \mathcal{A} \text{ and } s = 1, \ldots S, \pi_s^P(a) > 0 \text{ and } \pi_s^A(a) > 0. \tag{2.3}$$

As in GH, for simplicity we assume that common support condition (2.3) holds. Assumptions 1 and 2 in this model with heterogeneous beliefs ensure existence of an optimal first-best and a second-best contracts (For the details, see Lemma 3 in the appendix).

Finally, unless we specify otherwise, in what follows we will focus on the problem of finding the contract that implements a given action, rather than explicitly discussing the optimal action. As GH discuss, finding the optimal action is relatively straightforward.

3 Observable Actions

The first-best contract corresponds to the situation in which the agent's action is observable and the principal can explicitly make it part of the contract. In this case, the problem reduces to a standard risk sharing problem. When beliefs are identical, this problem has a well-known solution: pay the agent for the utility of the outside option plus the disutility of taking the desired action via a constant wage across states. This result follows from the simple observation that the (Pareto) optimal risk-sharing solution for a risk-neutral principal and risk averse agent entails the principal bearing all the risk. In our setting this result no-longer holds because (Pareto) optimal risk sharing between a risk-neutral principal and a risk-averse agent with heterogeneous beliefs will imply the latter still carries some risk.

When the agent's action is observable/contractible, the principal's contracting problem can be described as follows. For each action $a \in \mathcal{A}$, let w^a denote a contract so that the agent is willing to choose the action a: $U^A(w^a; a) = \sum_{s \in \mathcal{S}} \pi_s^A(a) w_s^a - g(a) \geq \underline{U}$. Over such a pair of actions and wage schemes the principal maximizes $U^P(w^a; a) = \sum_{s \in \mathcal{S}} \pi_s^P(a)[y_s - w_s^a]$. The principal's contracting problem can be divided into two parts: (1) for each action $a \in \mathcal{A}$ the principal finds the least costly way of implementing the action a; (2) given the cost of implementing each action and their expected output the principal optimally chooses the action a^* that maximizes the expected profit. Similar to GH, under Assumption 1 this problem can be simplified by a change of variables so that $v_s := u(w_s)$ and by taking inverse function $h := u^{-1}$, $w_s = h(v_s)$. The least costly way of implementing an observable action a solves the following minimization problem

$$\min_{v \in \mathbf{R}^S} \sum_{s=1}^{S} \pi_s^P(a) h(v_s) \tag{3.1}$$

subject to

$$\sum_{s=1}^{S} \pi_s^A(a) v_s - g(a) \geq \underline{u}. \tag{3.2}$$

Here, without loss of generality, we assume that if the agent is indifferent between two actions, then he will choose the one preferred by the principal. Under Assumptions 1 and 2, the first-best contract that solves this minimization problem exists (See part (a) of Lemma 3 in the appendix). When the principal and the agent have a common belief for each action $\pi^A(a) = \pi^P(a)$, GH show that the least costly way of implementing an action a is a fixed wage contract w^a independent of output realization that exactly compensates the agent for the effort cost: $u(w^a) - g(a) = \underline{u}$. This result does not necessarily hold under belief heterogeneity, as illustrated below.

In general, there may be more than one first-best optimal action and more than one first-best optimal incentive scheme.[2] However, in the cost minimization problem (3.1), the agent's utility function u is strictly concave and hence its inverse function h is strictly convex. Therefore, there is a unique first-best optimal incentive scheme which implements any particular first-best optimal action.

Notice also that in the cost minimization problem (3.1) the constraint set involves finite

[2]Here we focus on the characterization of the first best. We demonstrate the existence of an optimal first-best contract when beliefs are heterogeneous by generalizing GH's analysis of the *second*-best contract under a full common support assumption on the beliefs in part (a) of Lemma 3 in the Appendix.

number of linear conditions. Therefore, as there are finite number of actions in $\mathcal{A}$, the Kuhn-Tucker theorem implies that an optimal first-best contract is characterized as a solution to the necessary and sufficient first-order conditions for optimality:

$$\pi_s^P(a)h'(v_s) = \lambda\pi_s^A(a) \qquad \text{for all } s \in \{1,\ldots,S\}, \tag{3.3}$$

where the variable λ is the Lagrange multiplier corresponding to the individual rationality constraint (3.2). A simple evaluation of the equations (3.3) implies that an optimal contract must satisfy

$$\frac{\pi_s^P(a)h'(v_s)}{\pi_s^A(a)} = \frac{\pi_{s'}^P(a)h'(v_{s'})}{\pi_{s'}^A(a)} \qquad \text{for any } s, s' \in \{1,\ldots,S\}. \tag{3.4}$$

This condition characterizes Pareto optimal risk-sharing between individuals who have different beliefs. It can alternatively be rewritten as equality between the marginal rates of substitutions of the agent and the principal across any two output levels. Let $w^{FB}(a)$ be a contract that solves the principal's minimization problem (3.1) for implementing the action a, and denote by $C^{FB}(a;\pi^P(a),\pi^A(a))$ the expected cost for the principal, when she has belief $\pi^P(a)$ while the agent's belief is $\pi^A(a)$.

If the principal and the agent have common beliefs so that $\pi_s^P(a) = \pi_s^A(a) = \pi_s(a)$ for all $s \in \mathcal{S}$ and for all $a \in \mathcal{A}$, the optimality condition (3.4) implies that $h'(v_s)$ must be constant and independent of the output realization y_s. Hence, the value of the fixed wage w in the optimal contract is found by solving individual rationality (3.2) holding with equality. In this case, the first-best cost is $C^{FB}(a;\pi^P(a),\pi^A(a)) = C^{FB}(a;\pi(a),\pi(a)) = h(\underline{u}+g(a))$.

When the principal and the agent have different beliefs, i.e., $\pi_s^P(a) \neq \pi_s^A(a) = \pi_s(a)$ for at least two output levels, a constant payment w cannot satisfy the condition (3.4) and therefore the optimal first-best wage scheme cannot be constant. Given that the agent's participation constraints is independent of the principal's beliefs, the principal's minimization implies that $C^{FB}(a;\pi(a),\pi(a)) = h(\underline{u}+g(a)) \geq C^{FB}(a;\pi^P(a),\pi(a))$. The principal takes advantage of the disagreement with a contract that ends up being cheaper than $h(\underline{u}+g(a))$. [3]. Intuitively, when the principal is more optimistic than the agent about a state, the payment in that state will be lower; again, this is a simple consequence of Pareto efficient risk sharing. Since this is true irrespective of the output associated with that state, the first-best contract is not necessarily monotone in output in general. This analysis constitutes the proof of the following Lemma:

[3]As the agent's utility function u is strictly concave, and hence its inverse function h is strictly convex, The principal in most cases can achieve strictly smaller cost of implementation under belief heterogeneity so that $C^{FB}(a;\pi^P(a),\pi(a)) < C^{FB}(a;\pi(a),\pi(a)) = h(\underline{u}+g(a))$

Lemma 1. Under Assumptions 1 and 2, a first-best contract that implements an action a is generally not a constant wage when $\pi^P(a) \neq \pi^A(a) = \pi(a)$, unless $\pi^P(a) = \pi^A(a) = \pi(a)$, in which case the first best contract is constant. Additionally, the principal's cost of implementation is lower under belief heterogeneity: $C^{FB}(a; \pi(a), \pi(a)) = h(\underline{u} + g(a)) \geq C^{FB}(a; \pi^P(a), \pi(a))$.

As Lemma 1 shows the optimal first-best contract is typically non-constant. Moreover, we can also see from the optimality condition (3.4) that the optimal first-best contract is typically non-monotone, as for the action a, belief heterogeneity $\pi_s^P(a) \neq \pi_s^A(a)$ can take arbitrary forms. However, if the relative likelihood ratio in the optimality condition $\frac{\pi_s^A(a)}{\pi_s^P(a)}$ is monotone increasing(decreasing) in output levels y_s, then (3.4) implies that the first-best contract is monotone increasing(decreasing). We can therefore characterize monotonicity of the first-best contract if the beliefs of the principal and the agent are ordered according to the Monotone Likelihood Ratio Property (MLRP). [4]

We say that two probability distributions f and g over $\Delta(\mathcal{Y})$ are ranked according to MLRP, and f is said to dominate g in MLRP, if the relative likelihood $\frac{f(s)}{g(s)}$ is increasing in s. In that case, we write $f \succsim^{MLRP} g$. While it is sufficiently clear that MLRP ranking is an incomplete order of distributions over output levels, it is a useful order to classify belief heterogeneity as in the following definition:

Definition 1. We say that the agent is more optimistic about an action $a \in \mathcal{A}$ than the principal if $\pi^A(a) \succsim^{MLRP} \pi^P(a)$. Analogously, the principal is relatively more optimistic about an action a, if $\pi^P(a) \succsim^{MLRP} \pi^A(a)$.

MLRP ranking is familiar in the literature on moral hazard problems. In moral hazard problems, the parties have common beliefs about actions, and MLRP ranking over different actions is useful to characterize monotonicity of optimal contracts under unobservable actions. We show below that when the beliefs are heterogeneous, ranking of beliefs of the contracting parties for a given action characterizes monotonicity of optimal contracts when actions are contractible. Using Definition 1 and the first-best optimality condition (3.4), the optimal first-best contract that implements an action a is monotone increasing(decreasing) if the agent is more optimistic(pessimistic) than the principal. This observation is formalized in the following result.

Proposition 1. Consider an optimal first-best contract that implements an action a.

(a) If the agent is more optimistic about the action a than the principal, $\pi^A(a) \succsim^{MLRP}$

[4] Santos-Pinto (2008, Proposition 1) argues a similar result which relies on approximations and, hence, does not hold true generally. Our analytical result holds generally.

$\pi^P(a)$, then the first-best contract is monotone increasing in output.

(b) If the principal is more optimistic about the action a than the agent, $\pi^P(a) \succsim^{MLRP} \pi^A(a)$, then the first-best contract is monotone decreasing in output.

In addition to monotonicity, in this framework one can ask how the optimal contract responds to (possibly small) changes in the beliefs of one party. For instance, we consider perturbing the principal's beliefs so that the probabilities of only two states are different relative to the original distribution. In this case, the contract will reflect these changes in an intuitive way as stated in the following proposition.

Proposition 2. Consider $\pi^P(a)$ and $\widetilde{\pi}^P(a)$ such that: (i) $\pi_t^P(a) = \widetilde{\pi}_t^P(a)$ for all $t \neq s, s'$, (ii) $\pi_s^P(a) = \widetilde{\pi}_s^P(a) + \varepsilon$, and (iii) $\pi_{s'}^P(a) = \widetilde{\pi}_{s'}^P(a) - \varepsilon$, with $\min(\pi_s^P(a), \pi_{s'}^P(a)) > \varepsilon > 0$. Let $w^{FB}(a)$ and $\widetilde{w}^{FB}(a)$ be the first-best contracts corresponding to $\pi^P(a)$ and $\widetilde{\pi}^P(a)$, respectively. Then, $w_s^{FB}(a) \leq \widetilde{w}_s^{FB}(a)$ and $w_t^{FB}(a) \geq \widetilde{w}_t^{FB}(a)$ for all $t \neq s$ with at least two of the inequalities being strict.

This result shows that if the principal gives more weight to one state and less to some other state the contract must reflect this by paying the agent less in that state. Given that the agent's participation constraint is binding in any optimal contract, this implies that the principal should pay higher wages in other states. Proposition 2 illustrates how changes in beliefs lead to changes in the first-best contract. While the proposition is written in terms of changes to the principal's beliefs, one can obtain a similar result by perturbing the beliefs of the agent. Notice that it is straightforward to extend Proposition 2 to perturbations to beliefs that involve changes in multiple output levels.

3.1 First-Best Contract with CARA Utility

In what follows we exemplify the content of Proposition 2 by assuming that the agent's utility function belongs to the constant relative risk-aversion class, and there are only three states, $\mathcal{S} = \{1, 2, 3\}$. We assume the agent's Bernoulli utility function over monetary outcomes has the following form $u(x) = -e^{-x}$ where $x > 0$. For a wage level w, the agent's utility is given by: $-e^{-w}$.

In this setting, an interior solution solves the first-order conditions (3.3)

$$\pi_s^P(a) = \lambda \pi_s^A(a) e^{-w_s(a)} \qquad \text{for all } s \in \{1, \ldots, S\}, \tag{3.5}$$

together with the agent's participation constraint. The first-best optimal wage scheme (w_s^*)

follows directly by:

$$e^{-w_s^*(a)} = \frac{\pi_s^P(a)}{\pi_s^A(a)}(\underline{u} + g(a)). \tag{3.6}$$

Notice that consistent with Proposition 1 in the optimal risk-sharing contract the wage scheme satisfies: $w_s^*(a)$ is increasing in s if $\pi^A(a) \succsim^{MLRP} \pi^P(a)$; $w_s^*(a)$ is decreasing in s if $\pi^P(a) \succsim^{MLRP} \pi^A(a)$. Notice also the optimal $w_s^*(a)$ is decreasing in $\pi_s^P(a)$ and does not depend on $\pi_{s'}^A(a)$.

To illustrate the comparative statics of Proposition 2 in this setting, consider the ε reallocation between the probabilities of state 2 and state 3, i.e., $\pi_2^P(a) + \varepsilon$ and $\pi_3^P(a) - \varepsilon$. One can observe from (3.6) that as $\varepsilon > 0$ increases in the optimal risk-sharing contract: $w_2^*(a)$ decreases, $w_3^*(a)$ increases, while $w_1^*(a)$ remains unchanged. Hence, in the CARA utility case the comparative statics of perturbing beliefs shows that only the payments of the optimal contract for the states involved in the perturbation are affected.

4 Unobservable Actions

We now move on to analyzing the second-best situation in which the principal cannot observe the agent's behavior and therefore actions are not contractible. The optimal contract is such that the agent voluntarily chooses the action the principal would like to see implemented. In order to implement the desired action, the principal now has to deal with an incentive compatibility constraint in addition to the individual rationality constraint.

In what follows, we assume that a more costly action entails an "improvement" of both the principal's and the agent's beliefs in the sense of MLRP. Formally, this means that $g(a) > g(a')$ implies $\pi^P(a) \succsim^{MLRP} \pi^P(a')$ and $\pi^A(a) \succsim^{MLRP} \pi^A(a')$.

The principal's problem is to maximize her utility subject to individual rationality and incentive compatibility. This problem can be divided into two steps: (i) for any given action, find the optimal payment schedule that incentivizes the agent to take that action, and (ii) given these payment schedules, choose the action to incentivize. As in the previous section, we focus on the first of these problems as our attention is centered around optimal incentive schemes.

Similar to the standard case with homogenous beliefs, this problem is equivalent to minimizing expected wages, as expected revenues can be treated as a constant. In the first step, analogous to the problem (3.1), the principal solves,

$$\min_{v\in\mathbf{R}^S} \sum_{s=1}^{S} \pi_s^P(a)h(v_s) \tag{4.1}$$

subject to

$$\sum_{s=1}^{S} \pi_s^A(a)v_s - g(a) \geq \underline{u} \tag{4.2}$$

$$\sum_{s=1}^{S} \pi_s^A(a)v_s - g(a) \geq \sum_{s=1}^{S} \pi_s^A(a')v_s - g(a') \qquad \forall a' \in \mathcal{A}. \tag{4.3}$$

It is straightforward that the individual rationality constraint must hold as an equality; if not, the payments can be reduced by the same infinitesimal amount in all states without affecting the incentive compatibility constraints. In what follows, we analyze how heterogeneity in beliefs affects the characteristics of the optimal contracts.

4.1 Two Actions and Finite Number of Outputs

As we illustrate in the following analysis, analyzing the case of more than two actions available to the agent is complex. For that reasons, we restrict our setting so that only two actions are available to the agent: $\mathcal{A} = \{h, l\}$. The principal's cost minimization problem implies the following first-order conditions

$$\pi_s^P(h)h'(v_s) = \lambda\pi_s^A(h) + \mu\big[\pi_s^A(h) - \pi_s^A(l)\big] \qquad \text{for all } s \in \{1, \ldots, S\}, \tag{4.4}$$

where the terms λ and μ are the Lagrange multipliers corresponding to individual rationality and incentive compatibility constraints, respectively. Comparing with equation (3.3) one notices that, as usual, the first-best optimality condition is modified by the presence of an extra term that depends on incentives, reflected in the term with the Lagrange multiplier μ. This extra term involves a term that depends on the differences in the beliefs of the agent for the high and low action, $\pi^A(h)$ and $\pi^A(l)$, which is at the core of the incentive provision, weighted by the principal beliefs about the high action, $\pi^P(h)$. Notice that the principal's beliefs corresponding to the low effort action have no impact on the optimal contract that implements high effort (although they matter for the choice of which action to implement).

In this optimization problem, the Lagrange multipliers are strictly positive when the corresponding constraint binds. The multiplier λ cannot be zero because the individual rationality constraint must always bind. However, as illustrated below, in the case of two actions and two output levels, incentive compatibility constraint can be non-binding and, hence, the multiplier μ can be equal to zero at an optimum. The fact that incentive compatibility constraint can become irrelevant is a major difference with the case of common beliefs.

As in the case of the first-best, here we are interested in establishing conditions for the optimal contract to be monotone in output (i.e. increasing in y_s). The optimal second-best contract is characterized as a solution to the first-order conditions (4.4), individual rationality (4.2), and incentive compatibility (4.3). Notice that these conditions characterizing the set of optimal contracts do not depend on the probability distribution $\pi^P(l)$ of the low action perceived by the principal. This is a natural consequence of the fact that implementing the low action entails no agency cost to the principal.

We next turn to characterize the optimal contract under different scenarios consistent with the monotone likelihood ratio ranking. There are two possibilities depending on the relative ranking of the high action: (a) the agent has a more optimistic belief for the high action: $\pi^A(h) \succsim^{MLRP} \pi^P(h)$ and (b) the opposite, i.e., $\pi^P(h) \succsim^{MLRP} \pi^A(h)$.

Proposition 3. Let h be the action that is implemented by the principal in the optimal contract. (a) If $\pi^A(h) \succsim^{MLRP} \pi^P(h)$ then the wage scheme in the optimal contract is monotone increasing in output. (b) If $\pi^P(h) \succsim^{MLRP} \pi^A(h)$ then the wage scheme in the optimal contract can be non-monotone.

In Proposition 3 part (a) extends the monotonicity result of GH under the monotone likelihood ratio ranking to heterogeneous priors, provided the agent is more optimistic than the principal about higher output levels given that the high action is chosen. GH's Proposition 6 implies that in the case of two actions and common beliefs $\pi(h) \succsim^{MLRP} \pi(l)$ is sufficient to ensure that the optimal contract is monotone increasing in output. The result in part (a) therefore generalizes this result whenever the agent is more optimistic than the principal. part (b), more importantly, suggests that a failure of monotonicity is possible even if both parties find that the high action is better than the low action according to the MLRP ranking. When the agent is relatively more pessimistic than the principal, the MLRP condition, however, does not guarantee monotonicity. We illustrate this below with an example inspired by GH's Example 1.

Example 1. Consider an economic environment with two possible action $\mathcal{A} = \{h, l\}$ and three outcomes $\mathcal{S} = \{1, 2, 3\}$. The principal and the agent have the same beliefs about the low action: $\pi(l) := \pi^P(l) = \pi^A(l) = (\frac{2}{3}, \frac{1}{4}, \frac{1}{12})$, while their beliefs about the high action is given by: $\pi^A(h) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, $\pi^P(h) = (\frac{1}{12}, \frac{1}{4}, \frac{2}{3})$. Assume additive separability such that effort cost of the high action is $g(h) = (\frac{1}{12}\sqrt{\frac{5}{3}} + \frac{1}{4}\sqrt{\frac{7}{8}})$, $g(l) = 0$, and that the inverse of the agent's utility function is $h(v) = \frac{1}{3}v^3$. Notice that in this specification MLRP conditions holds for both the principal and the agent: $\pi^P(h) \succsim^{MLRP} \pi^A(h) \succsim^{MLRP} \pi(l)$.

We start with computing the cost of implementing the actions h and l, respectively, $C(h)$ and $C(l)$. It is clear that $C(l) = 0$. To compute the $C(h)$, we use the first-order conditions (4.4):

$$v_1^2 = 4\lambda - 4\mu, \quad v_2^2 = \frac{4}{3}\lambda + \frac{1}{3}\mu, \quad v_3^2 = \frac{1}{2}\lambda + \frac{3}{8}\mu,$$

together with individual rationality (4.2) and incentive compatibility (4.3) conditions. These conditions for the cost minimization problem of implementing action h are solved by setting $\lambda = \mu = 1$. This implies that $v_1 = 0, v_2 = \sqrt{5/3}, v_3 = \sqrt{7/8}$, and the agent is indifferent between h and l:

$$\frac{1}{3}v_1 + \frac{1}{3}v_2 + \frac{1}{3}v_3 - g(h) = \frac{2}{3}v_1 + \frac{1}{4}v_2 + \frac{1}{12}v_3$$

As the first-order conditions are both necessary and sufficient, we conclude that $C(h) = \frac{1}{4}\frac{v_2^3}{3} + \frac{2}{3}\frac{v_3^3}{3} = 0.36$.

Note that the incentive scheme with payments $w_s = \frac{1}{3}(v_s)^3$, which implements the action h,

$$w_1 = 0, w_2 = \frac{1}{3}(5/3)^{3/2}, w_3 = \frac{1}{3}(7/8)^{3/2}$$

is not non-decreasing: $w_1 < w_3 < w_2$.

Observe that $C(h) > C(l) = 0$. It is then easy to show that we can find finite output levels $y_1 < y_2 < y_3$ such that $\sum_s \pi_s^P(h)y_s - C(h) > \sum_s \pi_s(l)y_s$. This means that it is optimal for the principal to implement the action h using the incentive scheme $\{w_s\}$ described above. $\triangle$

In this example, the optimal contract is not increasing despite the fact that there are two actions and MLRP conditions are satisfied both for the principal and the agent. In the seminal work of GH with common beliefs, when there are two actions MLRP condition guarantees that the optimal contract is non-decreasing in output. Relative to GH, we apply the MLRP ranking to compare the heterogeneous beliefs over the same output levels, while each party's beliefs over different actions satisfy GH's MLRP condition separately. Even if we have such natural extension of MLRP conditions, the optimal contract fails to satisfy the same monotonicity property.

To understand the intuition behind this counter-example, it is key to realize that the principal's optimism results in higher payments in the states that the agent considers relatively more likely and, instead, lower payments in the states that the principal perceives more likely. Hence, differences in beliefs generate a gap that allows for optimal wage schemes that are non-monotone in output level.

GH in their Example 1 also show that in the common belief setting MLRP condition does not ensure monotonicity of optimal contracts when there are more than two actions. Yet, monotonicity is restored when a stringent condition, i.e.,concavity of distribution function condition (CDFC), is assumed in their framework. Our Example 1 shows that even when there are two actions, under belief heterogeneity MLRP condition for each party is not sufficient for monotonicity of the optimal contract. For that reason, it is not clear whether concavity can help in easing the analysis of the principal's problem. Extending the analysis beyond two actions is an open problem.

We have thereby shown that in the case of heterogeneous beliefs the classic results of GH do not apply in general. This is an important result since it highlights that in the presence of belief heterogeneity monotonicity is not always guaranteed. [5] Without more specific assumptions about functional forms of the utility function and about the probability distribution of the parties under the two actions, a sharper characterization of the optimal contract in the case in which the agent is relatively more pessimistic is not straightforward.

As in the case of observable actions we now turn to characterize how the optimal contract responds to changes in the beliefs of one party. In a similar vein as in Proposition 2, we consider changing the principal's beliefs to a nearby distribution in which the probabilities of only two states are different relative to the original distribution. For tractability, we use a parametric setting for the next result.

4.2 Two Actions and Two Outputs

Here, we focus on a two action and two output levels case and illustrate some of the major issues one faces when going to a more general setting. We denote by h and l the two actions available to the agent $\mathcal{A} = \{h, l\}$, where the cost of the action h is $g(h) = c$, and the action l has zero effort cost, $g(l) = 0$. We also assume that output levels are such that the principal always prefers to implement the high effort action h unless otherwise stated. Denote by $\pi_H^P(h)$ the principal's belief about the high output when action h is taken, while the agent's belief for this output is $\pi_H^A(h)$.

For simplicity, we assume that the principal and the agent have the same beliefs about the high output when the action l is taken, i.e., $\pi_H^A(l) = \pi_H^P(l)$. We also assume that higher cost action yields higher probability of high output both for the principal and the agent: $\pi_H^P(h) > \pi_H^P(l)$ and $\pi_H^A(h) > \pi_H^A(l)$. Notice that in the case of two output levels, a probability distribution over binary states is uniquely determined by the probability of high output.

[5] For example, in his analysis Santos-Pinto (2008) suggests erroneously that the monotonicity result holds in the case in which the agent is more pessimistic than the principal.

In this binary action and binary outcome case, the principal's cost minimization problem is given by:

$$\min_{\{w_L, w_H\}} \pi_H^P(h) w_H + (1 - \pi_H^P(h)) w_L$$

$$\text{s.t.}$$

$$\pi_H^A(h) u(w_H) + (1 - \pi_H^A(h)) u(w_L) - c \geq \underline{u} \tag{4.5}$$

$$\pi_H^A(h) u(w_H) + (1 - \pi_H^A(h)) u(w_L) - c \geq \pi_H^A(l) u(w_H) + (1 - \pi_H^A(l)) u(w_L) \tag{4.6}$$

The first-order conditions take the following form:

$$\pi_H^P(h) = [\lambda \pi_H^A(h) + \mu(\pi_H^A(h) - \pi_H^A(l))] u'(w_H) \tag{4.7}$$

$$1 - \pi_H^P(h) = [\lambda(1 - \pi_H^A(h)) - \mu(\pi_H^A(h) - \pi_H^A(l))] u'(w_L) \tag{4.8}$$

As mentioned earlier, in this setting, the individual rationality condition (4.5) always bind at an optimal contract andthe associated Lagrange multiplier $\lambda > 0$. Notice that the first order conditions (4.7) and (4.8) imply that at the optimum the Lagrange multiplier, μ, on the incentive compatibility condition (4.6) satisfies:

$$\mu \left(\pi_H^A(h) - \pi_H^A(l) \right) = \frac{(1 - \pi_H^A(h)) \pi_H^P(h)}{u'(w_H)} \left(1 - \frac{\pi_H^A(h)(1 - \pi_H^P(h))}{\pi_H^P(h)(1 - \pi_H^A(h))} \frac{u'(w_H)}{u'(w_L)} \right) \tag{4.9}$$

Clearly, the heterogeneity in beliefs determines how binding the incentive compatibility condition (4.6) can be. Notice that at an optimal contract when the incentive compatibility condition (4.6) binds so that $\mu > 0$ the optimal wages w_H and w_L solve both the individual rationality condition (4.5) and the incentive compatibility condition holding with equality and satisfy:

$$(\pi_H^A(h) - \pi_H^A(l))[u(w_H) - u(w_L)] = c \tag{4.10}$$

Here, since the agent assigns higher probability to the high output when he works rather than shirks (i.e., $\pi_H^A(h) > \pi_H^A(l)$), the optimal wage is increasing in output: $w_H > w_L$. As indicated in Proposition 3, this result is not general. This is because in the case of binary actions and output levels, the optimal contract for the two wage levels w_L and w_H are the unique solution to the individual rationality and the incentive compatibility constraints both holding with equality, which are entirely determined by the agent's beliefs. However, when the number of outcomes increases the principal's beliefs shapes the optimal wage scheme. To exemplify this, when there are two actions and three outputs, incentive compatibility and individual rationality constraints holding with equality gives a set of wage schemes that can implement the high action. This set typically contains more than one contract that can implement action h (two equations with three unknowns) and the principal uses his beliefs to chose the optimal among them.

Equation (4.9) implies that the Lagrange multiplier μ is strictly decreasing as the principal's belief $\pi_H^P(h)$ decreases. This results from using the fact that the optimal contract in the binary case is independent of the principal's beliefs when the IC binds and that $w_H > w_L$. Notice also that if the agent is relatively more optimistic about the high action so that $\pi_H^P(h) < \pi_H^A(h)$ and $\pi_H^P(h)$ is sufficiently low, then the Lagrange multiplier is nil, i.e., $\mu = 0$ and, hence, the incentive compatibility constraint does not bind. Given the agent's beliefs, the minimum of the principal's belief $\bar{\pi} > \pi_H^P(l)$ about the high action for which $\mu \geq 0$ is determined by

$$1 = \frac{\pi_H^A(h)(1-\bar{\pi})}{\bar{\pi}(1-\pi_H^A(h))} \frac{u'(w_H)}{u'(w_L)} \tag{4.11}$$

This therefore yields a two part characterization of the optimal contract depending on the belief heterogeneity. First, if the principal's belief about the high action is sufficiently more pessimistic compared to the agent's so that $\pi_H^P(h) \leq \bar{\pi}$, in the optimal contract that implements the high action, the incentive compatibility constraint does not bind and its allocation is the same as in the first-best contract. Naturally, when $\mu = 0$, the first-order conditions imply that the optimal wage is increasing(decreasing) in output if the agent is more optimistic(pessimistic) than the principal about the high action, as in case of observable action.

Second, if on the other hand the principal's belief is optimistic enough $\pi_H^P(h) > \bar{\pi}$ then the incentive compatibility constraint binds in addition to the individual rationality and the optimal contract differs from the first best. Moreover, interestingly enough, starting from the common belief case $\pi_H^P(h) = \pi_H^A(h)$, there is a range for the principal's belief $\pi_H^P(h) \in [\bar{\pi}, \pi_H^A(h)]$ for which the second-best contract is not responsive to changes in the principal's belief $\pi_H^P(h)$. Depending on the belief heterogeneity, the analysis therefore yields a two part characterization of the optimal contract that implements the High action and it is summarized in the following result:

Lemma 2. Consider an economic environment with two actions and two outputs, where the agent's belief about the High action is $\pi_H^A(h)$, the principal's belief about the high action is $\pi_H^P(h)$, and the parties have common belief $\pi_H^A(l)$ about the low action. If the principal's belief threshold $\bar{\pi}$ for the high action is given in (4.11), then in this setting the optimal contract that implements the high action satisfies:

(a) If $\pi_H^P(h) \leq \bar{\pi}$, then the optimal contract is the same as the first-best contract.

(b) If $\pi_H^P(h) > \bar{\pi}$, then the incentive compatibility constraint binds and the second-best contract differs from the first best. Moreover, in this case the optimal contract is determined only by the agent's beliefs and is not sensitive to the principal's belief about the high action.

Notice that in this Lemma, the threshold belief $\bar{\pi}$ is determined by the primitives as explained in the derivation of the equation (4.11).

In words, this result says that if the principal is not too optimistic about the high action, the optimal contract is monotone in output, and it is not sensitive to small changes in the principal's beliefs. On the other hand, when the principal becomes very pessimistic about it, the optimal contract for the high action is the first-best contract.

4.3 Comparative Statics: Two actions and Three output levels and CARA utility

In this section, we analyze how the optimal contract changes as disagreement between the principal and the agent increases by a small amount. Formally, the probability the principal assigns to some output level increases by some $\varepsilon > 0$, simultaneously the probability of a different output level is decreased by that same amount.

For simplicity, we consider an interior optimal contract that is a unique solution to the set of first-order conditions (4.4), when both individual rationality (4.2), and incentive compatibility (4.3) bind. The solution to this system of equations depends implicitly on the parameter ε. Formally, the wage scheme $(w_s^*(\varepsilon))_{\{s\in S\}}$ and the Langrange multipliers $\lambda^*(\varepsilon)$ and $\mu^*(\varepsilon)$ are functions of ε. Notice from (4.4) that the comparative static analysis on the optimal incentive scheme $(w_s^*(\varepsilon))_{\{s\in S\}}$ depends on how the Lagrange multipliers $\lambda^*(\varepsilon)$ and $\mu^*(\varepsilon)$ change as ε increases. Even for monotone contracts, i.e, $(w_s^*(\varepsilon))_{\{s\in S\}}$ increasing in s for any $\varepsilon > 0$, $\lambda^*(\varepsilon)$ and $\mu^*(\varepsilon)$ can be non-monotone in ε. This is the reason a tractable characterization of the comparative statics is not straightforward without making further assumptions on the contracting environment.

We employ the CARA utility example of section 3.1. That is, there are three possible outcome levels ($\mathcal{S} = \{1, 2, 3\}$) and the agent's utility function belongs to the CARA family. Given the CARA family assumption, there is a linear relationship between the utility and the marginal utility functions and can the optimal contract can be solved explicitly and conduct the comparative statics analysis as ε changes.

Proposition 4. Consider a moral hazard environment with two actions and three possible output levels. Moreover, the beliefs of the agent and principal satisfy MLRP conditions.

Assume the principal's probability distribution is such that one can write the probabilities of output levels y_s and $y_{s'}$ as $\pi_s^P(h)+\varepsilon$ and $\pi_{s'}^P(h)-\varepsilon$, for a small $\varepsilon>0$. Then, w_s is decreasing in ε and $w_{s'}$ is increasing in ε.

Remark 1: This result contrasts with the first-best comparative statics in the same environment with observable actions. In that environment, similar comparative static analysis shows that the incentives move in the direction of increasing disagreement (see Section 3.1). Recall that for the first best contract the wage was not affected in the output level whose probability unaffected (i.e., w_1^{FB}). Here, instead, the optimal second-best contract changes for the state that is not involved in comparative static analysis, in particular it increases (for the details, see the proof of Proposition 4 in the Appendix). Thus, the presence of moral hazard is reflected not only on the shape of the optimal contract, but also on the way it is affected by (small) changes in the beliefs of the principal.

In the appendix we show that tractable general comparative statics results analogous to Proposition 4 are difficult to obtain when one increases the number of outputs in this setting. In the two-by-three semi-parametric example analyzed above, we can explicitly characterize the solution for the optimal wage scheme in terms of the distributions perceived by the parties. However, as the number of output levels in the contracting problem increases, tractable characterization of the optimal contract is not straightforward. [6]

5 Conclusions

To what extent do the classic results that characterize optimal contracts in moral hazard settings hold up in an environment where the players think different? We show that allowing for belief heterogeneity opens the door to new trade-offs that can lead to contracts with significantly different properties.

Extending the GH framework with belief heterogeneity, we characterize the shape of the optimal contract and provide the literature with an important result: Contrary to the case of common beliefs, the MLRP rankings of the parties' beliefs do not imply that the wage scheme in the optimal contract is non-decreasing in output. Under MLRP condition, the monotonicity result of GH extends to belief heterogeneity when the agent is more optimistic than the principal. On the contrary, when the agent is relatively more pessimistic the optimal

[6] Following an iterative approach, we have solved for the optimal contract by iteratively increasing the number of outputs in the problem and show that this iterative method finds the solution to the optimal contract. However, it does not yield a tractable enough characterization of the optimal contract that could enable us to perform comparative statics analysis. This, again, is mainly due to the fact that the Lagrange multipliers $\lambda^*(\varepsilon)$ and $\mu^*(\varepsilon)$ can be non-monotone in ε. This derivations are available upon request by the authors.

contract can be non-monotone. This result is important since it implies that the MLRP condition of GH cannot be indiscriminately applied in the presence of heterogeneity of beliefs. We show that the MLRP condition in the case of two actions and finite output levels might not hold when beliefs are heterogeneous, refining the original results of GH.

We conclude by acknowledging that there are limitations to our analysis. First, we limit ourselves to only two actions and finitely many outputs, and for some results, we are restricted to only three output levels. In the two action case, we illustrate that the classical analysis on the properties of the optimal incentive schemes do not generalize to belief heterogeneity. What can be said about the properties of optimal contracts with arbitrary numbers of outputs and actions with reasonable degree of tractability is still very much an open question.

References

Borch, K. (1962). Equilibrium in a reinsurance market. *Econometrica: Journal of the Econometric Society*, 424–444.

De la Rosa, L. E. (2011). Overconfidence and moral hazard. *Games and Economic Behavior 73*(2), 429–451.

Dumav, M. and U. Khan (2021). Moral hazard, uncertain technologies, and linear contracts. *Working paper, Universidad Carlos III de Madrid*.

Fan, Y., A. Y. Orhun, and D. Turjeman (2020). Heterogeneous actions, beliefs, constraints and risk tolerance during the covid-19 pandemic. *Working paper, National Bureau of Economic Research*.

Gervais, S., J. B. Heaton, and T. Odean (2011). Overconfidence, compensation contracts, and capital budgeting. *The Journal of Finance 66*(5), 1735–1777.

Grossman, S. J. and O. D. Hart (1983). An analysis of the principal-agent problem. *Econometrica 51*(1), 7–46.

Holmström, B. (1979). Moral hazard and observability. *The Bell journal of economics*, 74–91.

Lopomo, G., L. Rigotti, and C. Shannon (2011). Knightian uncertainty and moral hazard. *Journal of Economic Theory 146*(3), 1148–1172.

Santos-Pinto, L. (2008). Positive self-image and incentives in organisations. *The Economic Journal 118*(531), 1315–1332.

A Appendix

Lemma 3. Under Assumptions1 and 2,

(a) there exists a first-best action and a first-best optimal incentive scheme;

(b) there exists a second-best action and a second-best optimal incentive scheme.

Proof of **Lemma 3**. The proof strategy extends and applies analogous arguments in the proof of GH's Proposition 1, which mainly exploits in the cost minimization problem to implement a given action, the convexity of the principal's objective function and the liner constraint set, to heterogeneous beliefs.

To establish the part (a), we show first, that, if the constraint set is nonempty for an action $a^* \in \mathcal{A}$, then the problem (3.1) has a solution, i.e. $\sum_s \pi_s^P(a^*)h(v_s)$ achieves its greatest lower bound $C(a^*)$. Note that $\sum_s \pi_s^A(a^*)v_s$ is bounded below by a finite number $\underline{U}$ on the constraint set of the problem (3.1). Therefore, applying analogous arguments as in the proof of Proposition 1 in GH, we can show that unbounded sequences in the constraint set implies that $\sum_s \pi_s^P(a^*)h(v_s)$ tend to infinity (roughly because the beliefs have a full common support (Assumption 2) the variance of $v_i \to \infty$ while their mean is bounded below, and h is convex and nonlinear. This in turn implies that the constraint set can be artificially bounded. As the constraint set is closed and the objective function is continuous, the existence of a minimum $C_{FB}(a)$ therefore follows from Weierstrass' theorem. Moreover, the cost function $C_{FB}(a)$ is lower semicontinuous in a.

Given, $C_{FB}(a)$ is lower semicontinuous and the action set $\mathcal{A}$ is compact, Weierstrass' theorem yields that $\max_{a\in\mathcal{A}}[B(a) - C_{FB}(a)]$ has a solution provided that $C_{FB}(a)$ is finite for some action a. To show this, if the principal and the agent have common belief, $\pi^A(a) = \pi^P(a)$ for all $a \in \mathcal{A}$, we denote the cost function by $\widehat{C}_{FB}(a)$. As in GH the lowest cost action a^*, which minimizes $\widehat{C}_{FB}(a)$, can be implemented by a constant wage scheme such that $u(w_s) = \underline{U} + g(a^*)$ for all $s \in \mathcal{S}$. As the fixed wage contract that implements the lowest cost action satisfies the constraint set in the problem (3.1) $C_{FB}(a^*) \leq \widehat{C}_{FB}(a^*)$. This therefore show that for some action a^* the constraint set is non-empty and the cost $\widehat{C}_{FB}(a^*)$ is finite and hence an optimal first-best action exists. This in turn establishes the existence of a first-best optimal incentive scheme, as we have also shown that if the constraint set is non-empty, the problem (3.1) has a solution.

We next turn to establish the part (b). As in the problem (4.1) the objective function is convex and the constraint set is linear, the problem in (4.1), following the arguments analogous to the proof of part a) yields the existence of a minimum $C(a)$ if the constraint set is non-empty. Moreover, the constraint set is non-empty as the incentive scheme which implements the action a^* in part a) satisfies the constraint set in the problem (4.1). The existence of a second-best action and an incentive scheme that implements it then follows analogously to the proof of part a).

□

Proof of **Proposition** 1. The result follows from the first-order condition of the first-best problem. For any two output levels y_s and $y_{s'}$ with $y_s > y_{s'}$ the first-order conditions (3.3) imply

$$\frac{\pi_s^P(a)}{\pi_{s'}^P(a)} = \frac{\pi_s^A(a)}{\pi_{s'}^A(a)} \frac{u'(w_s)}{u'(w_{s'})}.$$

This expression can be rearranged as

$$\frac{\pi_s^P(a)}{\pi_s^A(a)} = \frac{\pi_{s'}^P(a)}{\pi_{s'}^A(a)} \frac{u'(w_s)}{u'(w_{s'})}.$$

Note from this condition that if $\pi^P(a) \succsim^{MLRP} \pi^A(a)$ then $\frac{\pi_s^P(a)}{\pi_s^A(a)} \geq \frac{\pi_{s'}^P(a)}{\pi_{s'}^A(a)}$ and therefore the first order condition can only hold if $\frac{u'(w_s)}{u'(w_{s'})} \geq 1$ which implies $w_s \leq w_{s'}$ by the concavity of u. The proof of the other case follows analogous reasoning. □

Proof of **Proposition** 2. The proof uses the first order conditions and the concavity of the agent's utility function. The first order conditions in output y_s and any output level $y_t \neq y_s$ for the two different beliefs of the principal imply

$$\frac{\widetilde{\pi}_s^P(a) + \varepsilon}{\pi_t^P(a)} = \frac{\pi_s^A(a)}{\pi_t^A(a)} \frac{u'(w_s^{FB}(a))}{u'(w_t^{FB}(a))} \qquad \text{and} \qquad \frac{\widetilde{\pi}_s^P(a)}{\widetilde{\pi}_t^P(a)} = \frac{\pi_s^A(a)}{\pi_t^A(a)} \frac{u'(\widetilde{w}_s^{FB}(a))}{u'(\widetilde{w}_t^{FB}(a))}$$

Putting the two together one gets

$$\frac{\widetilde{\pi}_s^P(a) + \varepsilon}{\pi_t^P(a)} = \frac{\widetilde{\pi}_s^P(a)}{\widetilde{\pi}_t^P(a)} \frac{u'(w_s^{FB}(a))}{u'(w_t^{FB}(a)))} \frac{u'(\widetilde{w}_t^{FB}(a))}{u'(\widetilde{w}_s^{FB}(a))}$$

Since $\varepsilon > 0$ and $\pi_t^P(a) \leq \widetilde{\pi}_t^P(a)$ for all t, this last equality can only be satisfied if

$$\frac{u'(w_s^{FB}(a))}{u'(\widetilde{w}_s^{FB}(a)))} \frac{u'(\widetilde{w}_t^{FB}(a))}{u'(w_t^{FB}(a))} > 1$$

This shows that both fractions cannot be smaller than 1, and either $w_s^{FB}(a) \leq \widetilde{w}_s^{FB}(a)$, or $w_t^{FB}(a) \geq \widetilde{w}_t^{FB}(a)$ for all t, or both. Moreover, by individual rationality one cannot have that $w^{FB}(a)$ pays less than $\widetilde{w}^{FB}(a)$ in every output, while optimality implies that it cannot pay more in every output level. This establishes the result. $\square$

Proof of **Proposition** 3. To show part (a) notice that $\pi^A(h) \succsim^{MLRP} \pi^P(h)$ implies that likelihood ratios $\frac{\pi_s^A(h)}{\pi_s^P(h)}$ and $\frac{\left[\pi_s^A(h)-\pi_s^A(l)\right]}{\pi_s^P(h)}$ are both increasing in s. Since the utility function u is strictly concave and $\lambda > 0$ and $\mu \geq 0$, (4.4) implies that w_s is increasing in s.

We next to turn to show part (b). Notice that if $\pi^P(h) \succsim^{MLRP} \pi^A(h)$, then the likelihood ratio $\frac{\pi_s^A(h)}{\pi_s^P(h)}$is decreasing in s. However, the second term on the right-hand side of (4.4), $\frac{\left[\pi_s^A(h)-\pi_s^A(l)\right]}{\pi_s^P(h)}$, is not necessarily monotone decreasing. Therefore, whether the contract is monotone in this case is ambiguous. $\square$

Proof of **Proposition** 4. In this setting, for an interior solution $w_s > 0$ the first-order conditions take the following form:

$$\pi_s^P(h) = \left(\lambda\pi_s^A(h) + \mu(\pi_s^A(h) - \pi_s^A(l))\right) e^{-w_s}, \quad s \in \mathcal{S}. \tag{A.1}$$

The individual rationality constraint is

$$\sum_{s\in\mathcal{S}} \pi_s^A(l)(-e^{-w_s}) = \underline{u} + c \tag{A.2}$$

and the incentive compatibility

$$\sum_{s\in\mathcal{S}} (\pi_s^A(h) - \pi_s^A(l))(-e^{-w_s}) = c \tag{A.3}$$

Notice that the system of equations (A.1)-(A.3) is linear in marginal utilities e^{-w} and in the Lagrange multipliers, λ and μ. Using this system of equations we characterize the solution for the incentive scheme algebraically. It will be useful to simplify the notation and define the difference in the agent's beliefs by $\Delta_s := (\pi_s^A(h) - \pi_s^A(l))$.

We start by expressing the Lagrange multipliers in terms of the wages. Summing over the first-order conditions yields:

$$\lambda = \pi_1^P(h)e^{w_1} + \pi_2^P(h)e^{w_2} + \pi_3^P(h)e^{w_3} \tag{A.4}$$

and using the first-order condition for $s = 1$ we can write μ as follows:

$$\mu = \frac{\pi_1^P(h)}{\Delta_1}e^{w_1} - \frac{\lambda\pi_1^A(h)}{\Delta_1} \tag{A.5}$$

Here $\Delta_1 < 0$, since $\pi^A(h) \succsim^{MLRP} \pi^A(l)$ implies that $\frac{\pi_s^A(h)}{\pi_s^A(l)}$ is increasing in s. For otherwise, $\pi^A(h)$ and $\pi^A(l)$ would not be different distributions. For the same reason, $\Delta_3 > 0$.[7]

To simplify notation in what follows, let $\kappa_{s',s} := (\Delta_{s'}\pi_s^A(h) - \Delta_s \pi_{s'}^A(h)) = -\pi_{s'}^A(l)\pi_s^A(h) + \pi_s^A(l)\pi_{s'}^A(h) > 0$ whenever $s' > s$. Now we use the linear system of equations (A.2) and (A.3) to remove the terms with w_3:

$$\kappa_{3,1}e^{-w_1} + \kappa_{3,2}e^{-w_2} = (-\Delta_3 \underline{u} + c\pi_3(l)) \tag{A.6}$$

which can be solved to yield:

$$e^{w_2} = \frac{\kappa_{3,2}}{(-\Delta_3 \underline{u} + c\pi_3(l)) - \kappa_{3,1}e^{-w_1}} \tag{A.7}$$

Similarly, using (A.2) and (A.3) we express marginal utility at w_3 in terms of that at w_1:

$$\kappa_{2,1}e^{-w_1} - \kappa_{3,2}e^{-w_3} = (-\Delta_2 \underline{u} + c\pi_2(l)) \tag{A.8}$$

and solving:

$$e^{w_3} = \frac{\kappa_{3,2}}{\kappa_{2,1}e^{-w_1} - (-\Delta_2 \underline{u} + c\pi_2(l))} \tag{A.9}$$

Now we use an equation that has not been used so far to solve for the term w_1. For instance, we use the first-order condition (A.1) for $s = 2$:

$$\pi_2^P(h) = (\lambda \pi_2^A(h) + \mu \Delta_2)e^{-w_2}$$

Using (A.5) the latter implies:

$$\pi_2^P(h) = \left(\lambda \left(\pi_2^A(h) - \frac{\pi_1^A(h)\Delta_2}{\Delta_1}\right) + \left(\frac{\pi_1^P(h)e^{w_1}\Delta_2}{\Delta_1}\right)\right)e^{-w_2}$$

[7] More specially, the relationship $\pi^A(h) \succsim^{MLRP} \pi^A(l)$ implies:

$$\frac{\pi_1^A(h)}{\pi_1^A(l)} \leq \frac{\pi_2^A(h)}{\pi_2^A(l)} \leq \frac{\pi_3^A(h)}{\pi_3^A(l)}$$

and hence $\frac{\pi_1^A(h)}{\pi_1^A(l)} < 1$ and $\frac{\pi_3^A(h)}{\pi_3^A(l)} > 1$. For otherwise, if $\pi_1^A(l) \leq \pi_1^A(h)$, then $\pi_2^A(l) \leq \pi_2^A(h)$ and $\pi_3^A(l) \leq \pi_3^A(h)$. But this then implies that $\pi^A(h) = \pi^A(l)$. Therefore, $\Delta_1 < 0$ and $\Delta_3 > 0$.

Now, using (A.4) to substitute for λe^c in the last equation implies:

$$\pi_2^P(h) = \Bigg(\left(\pi_1^P(h)e^{w_1} + \pi_2^P(h)e^{w_2} + \pi_3^P(h)e^{w_3}\right) \underbrace{\left(\pi_2^A(h) - \frac{\pi_1^A(h)\Delta_2}{\Delta_1}\right)}_{:=\gamma_2} + \left(\frac{\pi_1^P(h)e^{w_1}\Delta_2}{\Delta_1}\right) \Bigg) e^{-w_2}$$

Simplifying and rearranging the previous equation implies:

$$\pi_2^P(h)(1-\gamma_2) = \left(\pi_1^P(h)e^{w_1}\left(\gamma_2 + \frac{\Delta_2}{\Delta_1}\right) + \pi_3^P(h)\gamma_2 e^{w_3}\right) e^{-w_2}$$

Finally, using (A.7) and (A.9) and substituting for e^{w_2} and e^{w_3}, respectively, yields:

$$\begin{aligned}\pi_2^P(h)(1-\gamma_2) = \Bigg(\pi_1^P(h)e^{w_1}\left(\gamma_2 + \frac{\Delta_2}{\Delta_1}\right) + \pi_3^P(h)\gamma_2 \frac{\kappa_{3,2}}{\kappa_{2,1}e^{-w_1} - (-\Delta_2 \underline{u} + c\pi_2(l))}\Bigg) \\ \left(\frac{(-\Delta_3 \underline{u} + c\pi_3(l)) - \kappa_{3,1}e^{-w_1}}{\kappa_{3,2}}\right)\end{aligned} \tag{M}$$

Here $\gamma_2 = \left(\pi_2^A(h) - \frac{\pi_1^A(h)\Delta_2}{\Delta_1}\right) = -\kappa_{2,1}/\Delta_1 > 0$ and $\gamma_2 \leq \pi_2^A(h) \leq 1$. Moreover, the MLRP relationship $\pi^A(h) \succsim^{MLRP} \pi^A(l)$ implies that $\left(\gamma_2 + \frac{\Delta_2}{\Delta_1}\right) > 0$, by an argument similar to that in the footnote 7.

The equation (M) (implicitly) determines the unique solution for w_1^* and relates it to the primitives in this setting. In turn, substituting for w_1^* in (A.6) and (A.8), respectively yields the unique solutions for w_2^* and w_3^*. Using (M) we can show that w_1 increases in ε. Suppose to the contrary that w_1 and hence e^{w_1} decreases as $\pi_2^P(h)$ increases. This implies that the left-hand side in (M) increases.

On the right-hand side, notice that the two terms in the product are both increasing in e^{w_1} (In the first term, larger e^{w_1} makes the denominator a smaller positive number and hence makes the ratio larger). Under the contradictory hypothesis, both the first and the second term on the right-hand side decrease. This in turn implies that starting with (M) holding with equality increasing $\varepsilon > 0$ leads to a decrease in the RHS while the LHS increases, which contradicts optimality. This therefore shows that the optimal w_1^* decreases in ε.

Now using the fact that w_1^* increases in ε implies, by (A.6) and (A.8), that the optimal w_2^* decreases and w_3^* increases in ε, respectively. This proves the claim in the statement of Proposition for states $s = 2$ and $s' = 3$. The comparative static result for any pair of outputs follows using the equation (M) and analogous reasoning. $\square$